

Investigating the effects of panel conditioning on the measurement of attitudes in the Life in Australia™ panel

D. Pennay, S. Slamowicz and P.J. Lavrakas

CSRM & SRC METHODS

PAPER NO. 2/2024

Series note

The ANU Centre for Social Research & Methods (CSRM) was established in 2015 to provide national leadership in the study of Australian society. CSRM has a strategic focus on:

- development of social research methods
- analysis of social issues and policy
- training in social science methods
- providing access to social scientific data.

CSRM publications are produced to enable widespread discussion and comment and are available for free download from the CSRM website (<https://csrm.cass.anu.edu.au/research/publications>).

CSRM is located within the Research School of Social Sciences in the College of Arts & Social Sciences at the Australian National University (ANU). The centre is a joint initiative between the Social Research Centre and the ANU. Our expertise includes quantitative, qualitative, and experimental research methodologies; public opinion and behaviour measurement; survey design; data collection and analysis; data archiving and management; and professional education in social research methods.

As with all CSRM publications, the views expressed in this Methods Paper are those of the authors and do not reflect any official CSRM position.

Professor Matthew Gray

Director, ANU Centre for Social Research & Methods
Research School of Social Sciences
College of Arts & Social Sciences
The Australian National University

March 2024

About the Social Research Centre

The Social Research Centre Pty Ltd provides research and evaluation services to Australia's social science research community in order to create new knowledge, inform decision-making and advance understanding of society. The Social Research Centre is owned by, and has deep links with, the Australian National University (ANU) and is co-founder of the ANU Centre for Social Research and Methods.

Authors

Darren Pennay is the Founder and former CEO of the Social Research Centre. He is an Honorary Professor in the Practice of Survey Methodology at the ANU Centre for Social Research & Methods.

Sam Slamowicz is a Survey Statistician at the Social Research Centre.

Paul J Lavrakas, PhD, is a Senior Methodological Adviser to the Social Research Centre, an Adjunct Professor at University of Illinois at Chicago, and a Senior Fellow with the National Opinion Research Center (NORC) at the University of Chicago.

Enquiries about this paper can be directed to darren.pennay@anu.edu.au.

Declaration of interests

This study was conducted and funded by the Social Research Centre. The probability-based online sample evaluated in this study – Life in Australia™ – is owned and operated by the Social Research Centre. This report expresses the views of the authors and not necessarily the views of the Social Research Centre.

Abstract

As evermore survey research is undertaken via online panels, including probability-based ones, an issue facing researchers is the possibility that data quality could be changed by panel conditioning effects that may arise from repeatedly interviewing the same respondents over time. This may especially occur when the same survey questions are asked repeatedly of the same panellists over data collection waves.

This paper presents findings from a study investigating the impact of panel conditioning on the reported attitudes of panellists in Australia's only probability-based online panel – Life in Australia™. The focus of this study is to look for changes in attitudinal reporting over time consistent with the theorised explanations provided by the Cognitive Stimulus Theory (CST) of Sturgis et al. (2009). CST hypothesises that repeated exposure to similar questions may lead to a non-negligible change in attitudes amongst some panellists, which will manifest as a crystallisation of expressed attitudes between the first wave of data collection and subsequent waves. This crystallisation of attitudes is hypothesised to lead, over time, to an increase in response validity and an increase in the inter-wave consistency of reported attitudes.

We find mixed results for the contention that panel conditioning leads to more valid responses being given over time. While we observe a statistically significant increase in the validity of responses for two of the five attitudinal constructs measured, two show no significant change and one shows a statistically significant decline in validity. We also find no clear

evidence that panellists' answers become more consistent as a result of repeatedly responding to the same questions over time. Across a battery of 33-items, panellists are almost as likely to demonstrate less consistency in their responses as more consistency. Even where we do see increases in response consistency this cannot be simply interpreted as support for the CST. We show that other factors such as the innate invariability of some of the attitudes being measured, as well as potentially harmful panel conditioning effects such as inter-wave freezing and satisficing, may also be contributing factors to the level of response invariance that we observe.

Another potentially beneficial effect arising from panel conditioning would be if there was a *reduction* in social desirability (SD) bias over time. This would accord with the theorised expectation that as panellists become more comfortable completing questionnaires, more familiar with repeated survey questions, and more confident and comfortable with the research panel environment, that they also become more willing to report less socially desirable behaviours and attitudes. However, we find evidence to the contrary and observe instead a small but non-negligible overall *increase* in socially desirable responding (SDR) across all 13 items we examine. The size of this effect is differential. When responding to a set of questions prone to SD bias, the amount of change observed is significantly different for various subgroups of panel members. The latent characteristics that seem to mediate changes in SDR over time include panellists' sociodemographic characteristics such as age and the socioeconomic status of their place of residence, personality traits relating to conscientiousness and agreeableness, as

well as the level of political activity engaged in. A panellist's starting out point in answering a set of items that are prone to SD bias is also a main predictor of the amount of subsequent change in the level of SDR. We put forward a possible alternative theoretical explanation, related to a sense of panel belongingness/attachment, as to why panel conditioning might bring about an increase rather than a decrease in SDR.

This study contributes to the emerging body of knowledge about panel conditioning effects in probability-based online panels and to the discussion as to whether such effects might be deemed beneficial or detrimental to data quality. Overall, we find little evidence in support of the Cognitive Stimulus Theory and, as such, little support for the assertion that panel conditioning has a beneficial impact on the quality of the attitudinal data reported by panellists when answering the same questions over various waves of a survey program.

Methods Paper No. 2/2024

An electronic publication downloaded from
<<http://csrcm.cass.anu.edu.au/research/publications>.

For a complete list of CSRM papers, see
<https://csrcm.cass.anu.edu.au/research/publications>.

**ANU Centre for Social Research
and Methods**

Research School of Social Sciences
College of Arts & Social Sciences
The Australian National University

Acknowledgments

The authors gratefully acknowledge the Scanlon Foundation Research Institute for making the data from their long-running *Mapping Social Cohesion* survey available for this study and for their ongoing support of the social sciences by routinely lodging their data with the Australian Data Archive.

We also thank the Social Research Centre for its ongoing commitment to methodological research which enabled this study.

A final note of thanks to the Life in Australia™ panellists for their ongoing participation in the panel.

Acronyms

ABS	Australian Bureau of Statistics
ANU	Australian National University
ATP	American Trends Panel
CATI	Computer Assisted Telephone Interviewing
CAWI	Computer Assisted Web Interviewing
CST	Cognitive Stimulus Theory
DFRDD	Dual-frame Random Digit Dialling
ESOMAR	European Society for Opinion and Market Research
SD	social desirability
SDR	socially desirable responding
SMS	Short Message Service (texting)

Contents

Acknowledgments	v
Acronyms	v
1 Introduction	2
2 Theoretical background and previous research	4
3 Methods	9
3.1 Data	9
3.2 Measures	10
3.2.1 Measuring changes in validity of attitudes over time	10
3.2.2 Measuring changes in the consistency of attitudes over time	11
3.2.3 Measuring changes in SDR over time	12
3.3 Significance testing	12
4 Results	14
4.1 Changes in the validity of attitudinal reporting over time	14
4.2 Changes in the inter-wave consistency of attitudinal reporting over time	15
4.3 Changes in SDR over time	19
5 Discussion	25
5.1 Our Research Questions	25
5.2 Strengths and limitations	26
6 Concluding remarks and next steps	28
References	29
Appendix 1 – Comparative profile of panellists	32
Appendix 2 – Substantive variables	34
Appendix 3 – Modelling	36
Overall model evaluation	38

Appendix 4 – Statistical Testing of inter-wave consistency 41

Tables

Table 1 Social cohesion domains and selected items..... 11

Table 2 The number of questions for which a respondent gave the same response in all four waves 17

Table 3 The number of times a panellist gave the same response to each 5-point Likert scale question (24 items) in all four waves 18

Table 4 Mean number of variables for which there was an increase or decrease in socially desirable responding 21

Table 5 Comparative profile of panellists by selected characteristics: Participated in fewer than four waves and participated in all four waves..... 32

Table 6 List of substantive variables and analytic use..... 34

Table 7 Model inputs 36

Table 8 Classification table..... 38

Table 9 Parameter Estimates – Increase in SDR between Waves 1 and 4..... 39

Table 10 Parameter Estimates – Decrease in SDR between Waves 1 and 4 40

Table 11 Gradients of the inter-wave measures of consistency and statistical significance testing 41

Figures

Figure 1 Optimal survey response process in a one-off (cross-sectional) survey (Tourangeau et al., 2000) 4

Figure 2 The Question Response Model for a longitudinal survey 5

Figure 3 Changes in the factor loadings for the for the items relating to the five domains over time 14

Figure 4 Inter-wave correlations: Consistency of the five factors over time 15

Figure 5 Inter-wave correlations: Inter-wave consistency of the 28 items over time 16

Figure 6 Means score for the 13 social desirability (SD) variables per wave 19

Figure 7 Proportion of panellists who showed a net increase, a decrease, or no change in social desirability (SD) across the 13 SD items 20

Figure 8 The relationship between a non-social desirability (SD) start point and increases in SDR over time 23

1 Introduction

Over the last two decades the dominant mode of data collection for polls and scientific surveys of the general population has changed from telephone interviews of persons sampled via randomly generated lists of telephone numbers (mobile and/or landline numbers), to questionnaires administered via the internet – so called, Computer Assisted Web Interviewing (CAWI, online) – to persons recruited into research panels, where all or almost all panellists provide data via CAWI. The European Society for Opinion and Market Research (ESOMAR) published the 2022 Global Market Research Report that shows that 4.4 per cent of global market research industry revenue is derived from data collected via CATI surveys compared with 32.6 per cent from online surveys (ESOMAR, 2022, 163). Most of these online surveys are conducted via research panels that use non-probability sampling methods to build and maintain their pool of ready-made survey participants but, as noted by Kennedy and her colleagues (Kennedy et al., 2023), in the United States at least (and elsewhere) increased use is being made of probability-based online panels.¹

One of the features of both probability and non-probability online panels is that panellists who remain active in such

panels will complete many questionnaires over time. This raises the prospect that panel conditioning effects could increase the amount of respondent-related measurement error in the data provided.² Panel conditioning is hypothesised to change the way that some panellists complete repeated survey questions and perhaps even change the actual and/or reported attitudes and opinions of panellists. Kraemer et al. (2023, 1) describe the learning effects arising from panel conditioning ‘as a major threat to response quality in the later waves of a panel study’.

It is from within this context that we look for evidence of panel conditioning in the reported attitudes of panellists in the probability-based research panel – Life in Australia™ – owned and operated by the Social Research Centre. The potential for panel conditioning to affect the attitudes reported by panellists is a particular concern given that Life in Australia™ hosts several longitudinal and time series surveys and given that it is the only probability-based research panel in Australia. The research findings reported in this paper are part of a series of planned investigations into the quality of the data provided to, and the accuracy of the estimates produced from, questionnaires administered on Life in Australia™.

This paper looks at the reported attitudes of panellists using data from the Scanlon

¹ Bracketed term added by the authors.

² Panel conditioning is an effect sometimes seen in repeated surveys when a panellist’s response is influenced by prior interviews or contacts. For example, in surveys that ask for

opinions, attitudes or projected behavior, a panellist may become more aware or informed of the issues simply through participating in a series of interviews (Cantwell, 2008, 566–567).

Foundation Research Institute's³ long-running *Mapping Social Cohesion Survey*. The survey commenced in 2007 and has been administered on Life in Australia™ since 2018.

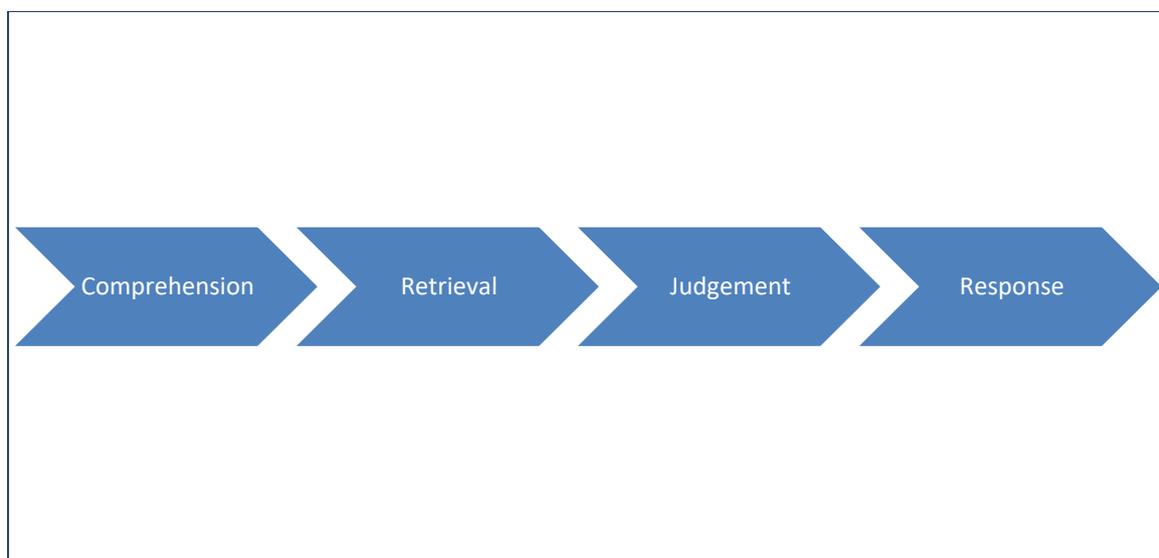
³ See <https://scanloninstitute.org.au/>.

2 Theoretical background and previous research

As shown in Figure 1, the theoretical survey response model established by

Tourangeau et al. (2000, 7–16) identifies that to give the optimal response to a survey question a respondent should go through at least four consecutive cognitive processes – (a) comprehension of the item, (b) retrieval of the relevant information from memory, (c) use of that information to make the required judgements, and (d) selection and reporting of an answer.

Figure 1 Optimal survey response process in a one-off (cross-sectional) survey (Tourangeau et al., 2000)



The survey response process described above applies when respondents are exposed to a questionnaire *just once*, such as in a standalone cross-sectional survey.

However, for formally constituted longitudinal surveys (i.e., when a respondent knows that they will be periodically re-interviewed on the same topic) and for repeated cross-sectional surveys on online panels (when a respondent may be re-interviewed but has no explicit prior knowledge of such), our interest is in understanding what impact, if any, this *repeated exposure* to the same survey questions may have on the question response model and the subsequent quality of the data provided.

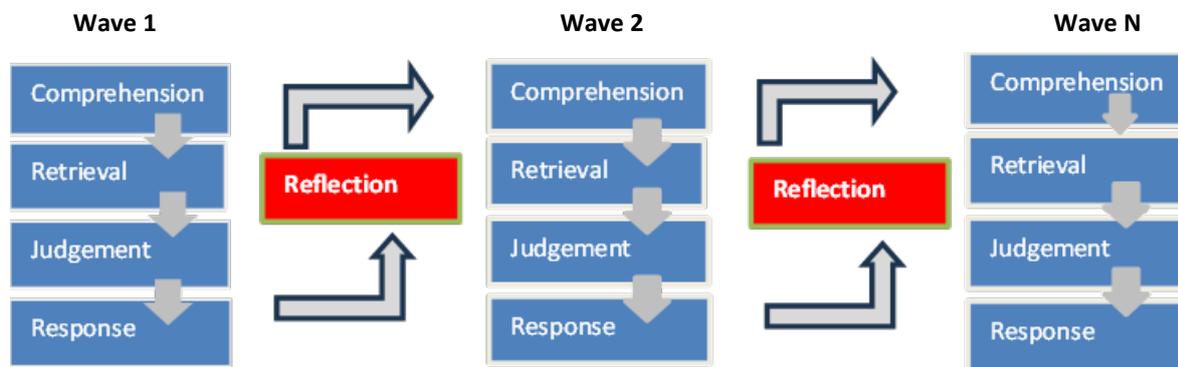
A theoretical mechanism used for explaining the impact of repeated questioning on actual and reported attitudes is the Cognitive Stimulus Theory (CST) (Sturgis et al., 2009, 116–117).

As discussed by Struminskaya and Bosnjak (2021, 279–280) the CST works on the premise that respondents who are prompted by repeatedly completing the same questions across waves of data collection will sometimes be motivated to gather information about interview topics, *between waves*. This may result in a panellist/respondent forming opinions where none existed before and/or in changing pre-existing opinions. As a result, the CST posits that attitudes will

become more consistent (i.e., crystallised) over the waves of a panel and attitude strength is predicted to increase. For example, after an initial interview, respondents may discuss the issues raised in that interview with friends and family, deliberate privately, and/or acquire new information about the issues by paying closer attention to the news media. The hypothesised result of engaging in these

behaviours is that respondents arrive at rational preference-based judgements as opposed to attitudes constructed ‘on the spot’⁴ (Sturgis et al., 2009, 279). This additional rational response process of ‘reflection’ could take place on one or more occasions following participation in a given survey wave. A possible survey response process which allows for inter-wave reflection is shown in Figure 2.

Figure 2 The Question Response Model for a longitudinal survey



Kraemer et al. (2023) distil three main panel conditioning mechanisms from the literature; (1) reflection, (2) satisficing⁵ and (3) social desirability.⁶ Our interest in this paper is the effect of panel conditioning on the reported attitudes of respondents, not on the process of

questionnaire completion so, to this end, we explore our data for evidence of changes in attitudes across waves due to reflection and social desirability (SD).⁷ The changes in attitudes that are hypothesised to occur as a result of reflection are:

⁴ Attitude formation ‘on the spot’ is not meant to suggest that one’s intuition is what is underlying the attitude.
⁵ Satisficing occurs when respondents ‘choose to perform one of more of the steps [in the Tourangeau et al. model] in a cursory fashion, or they may skip one or more of the steps altogether’ (Anand, 2008, 797–799).
⁶ ‘Social desirability bias is the tendency of some respondents to report an answer in a way that they deem to be more socially acceptable than would be their “true” answer.

They do this to project a more favourable image. The outcome of the bias is an overreporting of socially desirable behaviors or attitudes and an underreporting of socially undesirable behaviors or attitudes. Social desirability bias intervenes at the last stage of the Tourangeau et al. response process when the response is communicated to the researcher’ (Callegaro, 2008, 825–826).
⁷ Satisficing behaviours among Life in Australia™ panel members will be explored in another paper.

(1) attitudes may become more valid⁸
(2) attitudes may become more consistent; and (3) the degree of opinionation on issues may increase, that is, the use of ‘don’t know/can’t say’ response options will reduce over time (Sturgis et al., 2009). Given the data we have at our disposal we can explore only the first two of these outcomes.

With respect to the impact of repeated interviewing on SD bias, the thinking is as follows: ‘repeated interviewing can reduce social desirability bias by way of respondents learning the rules of the interview procedure. With each successive wave of interviewing, respondents may develop more trust in the survey organisation and the survey procedure, which, in turn, increases respondents’ trust and willingness to disclose unflattering information’ (Waterton & Lievesley, 1989, as summarised by Struminskaya & Bosnjak 2021, 285).

While Waterton and Lievesley provide empirical support for this hypothesis, as does Brannen (1993), no such effects are found by Mavletova and Lynn (2019) or Pevalin (2000), and the opposite effect is observed by Kraemer et al. (2023) and Torche et al. (2012). So, while there is empirical support for the theoretical view that panel conditioning can lead to a reduction in SDR, the findings on this topic are mixed and seemingly dependent upon the survey context – i.e., the perceived sensitivity of the questions and

respondent characteristics and how these interact. In addition, the older and possibly dated research that finds evidence of panel conditioning effects consistent with the CST concept of ‘reflecting’ on answers to questions between waves, mainly comes from large-scale longitudinal surveys such as the British Social Attitudes Survey (Waterton & Lievesley, 1989), the British Household Panel Survey (Sturgis et al., 2009), the Survey of Income and Program Participation (McCormick et al., 1992; Pennell & Lepkowski, 1992), the Current Population Survey (e.g., Bailar, 1975; Halpern-Manners & Warren, 2012) and the Health and Retirement Study, (Weir & Smith, 2007; Wilson & Howell, 2005, 2007). What distinguishes these large-scale, often face-to-face, longitudinal surveys from repeated cross-sectional surveys administered on probability-based research panels, is that for longitudinal surveys the respondents know that they will be re-interviewed whereas, more typically, when members of online research panels are exposed to repeated cross-sectional surveys, they are not informed that they may be re-interviewed on the same topic in a subsequent wave.

The expectation, or lack thereof, that one is going to be re-interviewed on a topic, may change how the process of reflection applies to these repeated interview situations. It seems reasonable to assume that, overall, panellists may be more prone to reflect on previous answers if they are explicitly informed that they will

⁸ In this context the concept of ‘validity’ means that the individual attitudinal items better measure the underlying constructs to which they relate. This is measured by the ‘intercorrelation among the scores on all

items in the test designed to measure the same thing. This type of estimate reflects content-sampling error and chance-response tendencies. It is a measure of consistency or homogeneity’ (Broedling, 1974, 374–375).

be re-interviewed on that topic, as is the case for formally constituted longitudinal surveys, compared to panellists who are not informed about the prospect of being re-interviewed, as is the case for repeated cross-sectional surveys. As such, inter-wave reflection may have more of an influence on the responses given in subsequent waves of a longitudinal survey than in subsequent waves of a repeated cross-sectional survey.

The available evidence appears to support this assertion. There are only a few studies that have looked for evidence of panel conditioning in probability-based online panels. Kraemer et al. (2023) in their study looking at evidence of panel conditioning in the GESIS panel⁹ do not find any evidence of content learning effects which would manifest as a decrease in the proportion of Don't Know responses across three cohorts with different levels of panel experience (Kraemer et al. 2023, 15). Nor do they find any evidence of a decrease in socially desirable response patterns (when their expectation was that a decrease in SD bias would result in a reduced prevalence of item non-response to a range of sensitive questions when comparing two cohorts with different levels of survey experience). In fact, they found the opposite (Kraemer et al., 2022, 18).

Similarly, no evidence of panel conditioning effects was found by Amaya et al. (2022) when looking for evidence of such effects due to reflection or SD in the American Trends Panel (ATP).¹⁰ Specifically, they found 'no evidence that conditioning has biased ATP estimates for news consumption, discussing politics, political partisanship or voting, though empanelment led to a slight uptick in [reported] voter registration' (Amaya et al., 2022, 3).

Finally, research into the Knowledge Panel¹¹ failed to find systematic conditioning effects on measures such as media consumption, interest in politics and public affairs and candidate preference (Clinton, 2001, 21–31).

So, it seems that the underlying question persists: Does repeated exposure to the same or similar questions on a probability-based online panel change the actual and/or reported attitudes of panellists?

The research questions we address in this paper are as follows:

Research Question 1 – Will we observe an increase, decrease or no change in the validity of attitudinal responses from panellists to the same items over time?

Research Question 2 – Will we observe an increase, decrease or no change in the

⁹ The GESIS Panel, established in February 2014, is part of [GESIS](#), Leibniz Institute for the Social Sciences in Mannheim, Germany. It offers the social science community an opportunity to collect survey data within a probability-based mixed-mode panel. Waves are fielded every three months with questionnaire length averaging 20–25 minutes to complete.

¹⁰ Created in 2014, the ATP is Pew Research Center's nationally representative online survey panel of the United States. The panel is composed of more than 10,000 adults selected at random from across the entire nation.

¹¹ The Knowledge Panel was the first probability-based panel established in the United States by Knowledge Networks in 1999.

consistency of responses from panellists
to the same items over time?

Research Question 3 – Will we observe an
increase, decrease or no change in socially
desirable responding to the same set of
sensitive items over time?

3 Methods

3.1 Data

As mentioned previously, our data come from the Mapping Social Cohesion Survey and were made available for the research we report here by the Scanlon Foundation Research Institute.

Commencing in 2007 and until 2012, the cross-sectional survey was administered by undertaking telephone interviews with adult Australians sampled from randomly generated landline telephone numbers. From 2013 to 2019, landline and mobile phone numbers were sampled (known as Dual-frame Random Digit Dialling, DFRDD). In 2018 and 2019, the survey was undertaken both via DFRDD and on the Social Research Centre's Life in Australia™ panel. Since 2020, the survey has been undertaken exclusively on Life in Australia™.

Life in Australia™ was established in 2016 and is Australia's first and still remains the country's only national probability-based online panel.¹² The panel commenced with 3,042 panellists being recruited, and since then, some panel members have left the panel or been retired and new panel members recruited using a variety of methods including DFRDD, via residential

addresses drawn from the Geocoded National Address file, and SMS Push to Web methods administered to randomly generated mobile phone numbers. At the time of writing, Life in Australia™ has 10,099 active members. Panel members are paid \$20 to join the panel and a further incentive of \$10 for each questionnaire completed, paid by gift voucher, deposit into a PayPal account, or charitable donation. Life in Australia™ includes people both with and without internet access. Those without internet access and those who are not comfortable completing questionnaires online are able to complete questionnaires via a telephone interview. Generally, approximately 3–5 per cent of panellists opt to complete any given wave via a telephone interview.

The data used for this study are from the July 2018, July 2019, July 2020, and July 2021 waves of the Mapping Social Cohesion survey. These surveys were administered in Waves 18, 29, 40 and 51 of Life in Australia™ and undertaken as standalone surveys, not as part of an omnibus questionnaire.¹³ The questionnaires took an average of 15–20 minutes to complete. At the core of each questionnaire is a range of attitudinal items relating to five domains of social cohesion (Sense of Belonging, Sense of Worth, Social Inclusion and Justice,

¹² While called an 'online' panel throughout this report, Life in Australia™ uses multiple modes of data collection with around 3%–5% of completed questionnaires per wave (the offline proportion of the panel), being gathered via CATI.

¹³ A November 2021 wave was undertaken to monitor the impact of the COVID-19 pandemic and related public health measures on social cohesion. While all the panellists in our

sample participated in the November 2021 wave, it has been excluded from our analysis so to maintain four equidistant 12-monthly intervals between waves. This is particularly important for the examination of response consistency over time where one could safely assume, and our data showed, greater consistency in the responses given when the time between waves was reduced.

Participation and Acceptance: Rejection). For full methodological details about Life in Australia™ see Kaczmirek et al., (2019) and Phillips et al., (2022).

The variables used for this study are a subset drawn from the responses to the more than 240 unique attitudinal variables administered to 4,519 respondents across these four waves of the survey. The data we use for our analyses comprises responses to 28 attitudinal and behavioural questions asked in each wave, along with a comprehensive range of profiling variables, available for the 1,333 panellists who responded to each of the four waves (see Appendix 1 for a profile of panellists and Appendix 2 for a complete list of the substantive variables included in our data set). Of the 1,333 panellists who completed all four waves of the survey, 1,251 (94%) were recruited at the beginning of the Life in Australia™ panel and 82 (6%) were recruited at Wave 18. Those recruited as Wave 18 were part of a small and very targeted top-up exercise undertaken by the panel managers which involved the use of CATI to recruit English-speaking Australian residents aged 18–54 years via a randomly generated mobile phone sampling frame. The decision was taken to keep these top-up panellists in our analytical sample because they help to boost the sample for some low count subgroups (e.g., 18–34-year-olds) and their inclusion does not change any of our substantive findings. For all intents and purposes this is a single cohort study.¹⁴

Due to the effects of panel attrition and wave-on-wave non-response, those panellists who completed all four waves of the survey differ from those who did not. Reference to Appendix 1 shows that relative to the population of 4-wave non-completers, 4-wave completers are significantly more likely to be male; aged over 54 years; have at least a Bachelor's Degree level of educational attainment; be overseas born from an English-speaking background; be currently married; be more frequent users of the internet; have voted for a conservative party at the last federal election; and report higher levels of psychological distress.

Despite these differences, an advantage of limiting our sample to those panellists who completed all four waves is that doing so means that our analyses are confined to the same respondents at each wave thereby avoiding wave on wave differences in the sample composition that may otherwise occur due to differential inter-wave unit non-response and attrition.

3.2 Measures

3.2.1 Measuring changes in validity of attitudes over time

The data available to us for this study contain two items, from a larger subset of items, relating to each of the five psychometrically constructed¹⁵ domains of social cohesion. This provides us with 10 questions, grouped into five pairs, for

¹⁴ Analysis was undertaken to ensure that this was the case.

¹⁵ Psychometrics originally comes from the field of measurement psychology. When applied to

survey research, psychometric tests are applied to determine how well a series of questions measure the underlying construct of interest.

which we can determine changes in how well each pair of questions relate to their underlying domain/construct and whether this changes over time, that is, we can measure changes in attitude validity over time. For our results to be consistent with the CST, we would need to observe an increase in the strength of the correlation between the question pairs and their factors over time.

The approach used is as per Sturgis et al. (2009) and involves five confirmatory factor models for item groupings defined by our Social Cohesion domains, each with one underlying factor. For each factor model, the communalities of associated

variables range from 0 to 1 and are interpreted as the proportion of variance of each variable explained by the underlying factor. By calculating the mean of communalities, we have the average proportion of variance of items which is explained by the underlying factor. This validity measure was calculated for each of the paired sets for all four waves showing how well the question pairs in each wave measure the intended underlying factor.

The items used for this analysis and their respective domains are provided in Table 1.

Table 1 Social cohesion domains and selected items

Domain	Items
Sense of Belonging	Extent you take pride in Australian way of life. Extent you have sense of belonging.
Sense of Worth	How satisfied with financial situation. Over the last year have you been happy / unhappy.
Social Inclusion and Justice	People on low incomes get enough government financial support. Australia is a land of economic opportunity
Participation	Last 3 years or so: Written or spoken to a Federal or State Member of Parliament. Last 3 years or so: Joined a boycott of a product or company.
Acceptance: Rejection	Accepting immigrants from many different countries makes Australia stronger. Ethnic minorities in Australia should be given Australian government assistance to maintain their customs and traditions.

3.2.2 Measuring changes in the consistency of attitudes over time

To measure any change in the inter-wave consistency of the attitudes being reported over time we summed the scales from confirmatory factor analysis models

(as described above), along with 28 other attitudinal items, and used either Pearson's or Spearman's correlation coefficient, as appropriate, to measure changes in the level of correlation between wave-to-wave responses to the same items over time. This was calculated

for three adjacent wave pairings: July 2018:July 2019, July 2019:July 2020 and July 2020:July 2021. The items used in this analysis are provided in Appendix 2.

3.2.3 Measuring changes in SDR over time

From the 28 attitudinal and behavioural questions asked in each wave we identified 13 items reasoned to be the most susceptible to SD bias. These items measure attitudes to immigration (8 items), distribution of wealth/equity (2 items) and discrimination (3 items), see Appendix 2.

All the items have strong face validity¹⁶ and external validity¹⁷ was supported by the finding that all items showed a marked decrease in SD (i.e., the more socially desirable response options were chosen less often) when the questionnaire was administered on the 2018 and 2019 Life in Australia™ panel compared to the parallel DFRDD surveys in 2018 and 2019. This is consistent with a reduction in SD bias because of moving from the interviewer-administered DFRDD telephone survey to the almost exclusively self-administered online completion mode used by Life in Australia™. This supports the use of these items as ones that are most likely to be susceptible to SD bias.

For the purposes of our analysis we define a change in SDR as having occurred when a panellist changes their reported opinion from one side of a proposition to another. For example, from disagreeing ‘that accepting immigrants from many different countries makes Australia stronger’ to agreeing with this proposition. This definition is used rather than treating any change in the level of agreement/disagreement with a proposition as an indicator of a change in SDR, as it provides a more robust indication that a change in the underlying SD of the response has occurred.

Two metrics are used to quantify the amount of change in SDR over time: (1) Change of direction – whether or not there has been a net increase, decrease or no change in SDR across these 13 items when comparing results from the 2018 and 2021 surveys, and (2) Change of magnitude – the average number of questions for which each panellist showed either an increase or decrease in SDR as reflected by their responses to these questions.

3.3 Significance testing

Unless otherwise stated, where results have been tested for statistical significance, this has been done using a bootstrap approach, due to the non-parametric nature of measures. The bootstrap approach is a re-sampling

¹⁶ Face validity is about whether a variable or scale appears to measure what it is supposed to measure. This type of validity is concerned with whether, on the surface, a measure seems relevant and appropriate for what it is meant to assess.
<https://www.scribbr.com/methodology/face-validity/>

¹⁷ External validity is the extent to which you can generalise the findings of a study to other situations, people, settings, and measures. In other words, can you confidently apply the findings of your study to a broader context?
<https://www.scribbr.com/methodology/external-validity/>

method, founded on the principle of re-sampling from observed data to simulate multiple iterations of the observed survey. To assess the significance of results, 5,000 re-samples of the responding population are used.

p-values therefore represent the estimated probability of the observed result happening by chance. More details of this bootstrap approach can be found in Davison et al. 1997.

4 Results

4.1 Changes in the validity of attitudinal reporting over time

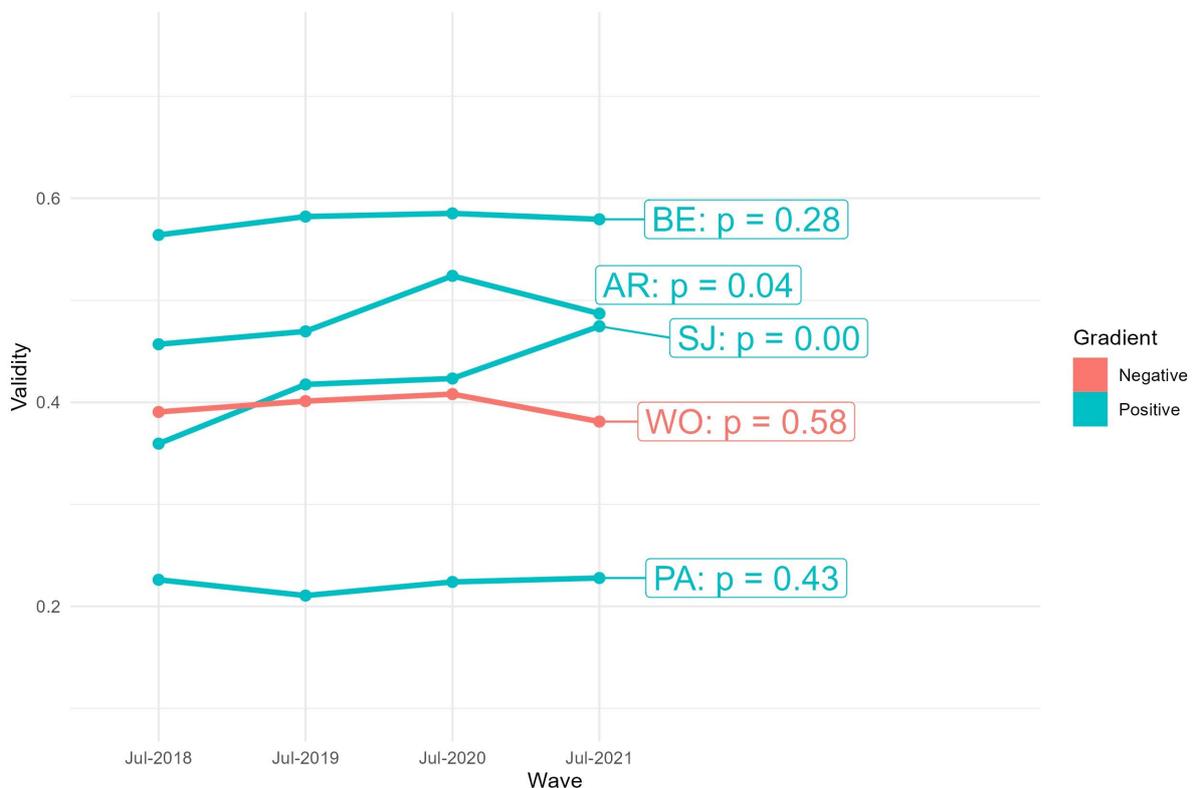
The reader should note that to assess the significance of the direction of change in the figures presented in this section, the plotted gradients are estimated using linear least squares estimation. The p-values in Figure 3, for example, represent the significance of the direction of the gradient, either positive or negative.

Our first research question is whether there is an increase in the validity of

attitudinal reporting across waves as predicted by the CST. The starting point is the average factor loadings for each item pair in July 2018, and then for each of the three waves thereafter. Two of the five factors (Social Justice and Acceptance: Rejection) show a statistically significant increase in validity, two (Belonging and Participation) show no significant change and the Worth domain shows a statistically significant decline in validity, as measured.

These mixed results provide insufficient evidence to support the CST hypothesis that panel conditioning is associated with an increase in the validity of attitudinal reporting over time.

Figure 3 Changes in the factor loadings for the for the items relating to the five domains over time



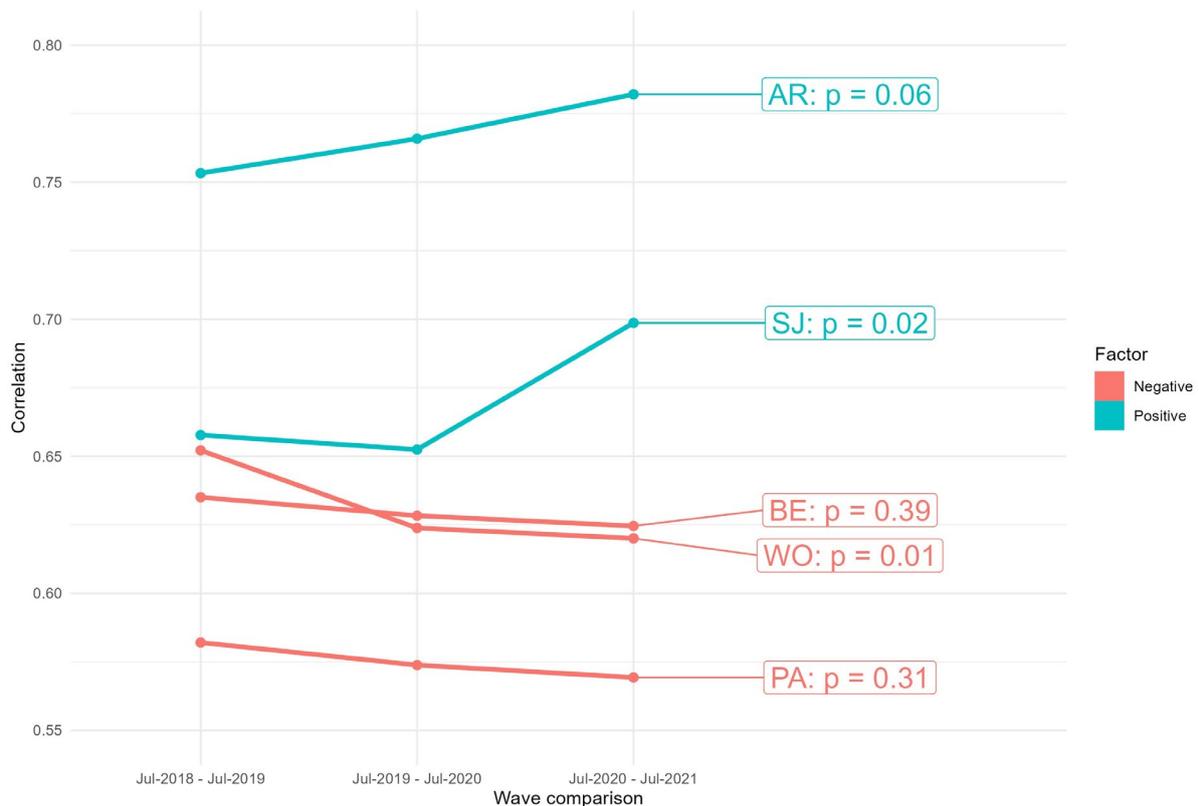
The p-values represent significance of direction of the estimated gradient, either positive or negative.

4.2 Changes in the inter-wave consistency of attitudinal reporting over time

Our second research question is whether the increase in the inter-wave consistency of attitudinal reporting over time, as predicted by the CST, is present in our

data. Figure 4 shows the changes in the mean inter-wave correlations for the five pairs of items used in our analysis of validity. The same two item pairs that showed a significant increase in validity – Social Justice and Acceptance: Rejection – also show significantly higher inter-wave correlations over time (2018 compared with 2021).

Figure 4 Inter-wave correlations: Consistency of the five factors over time

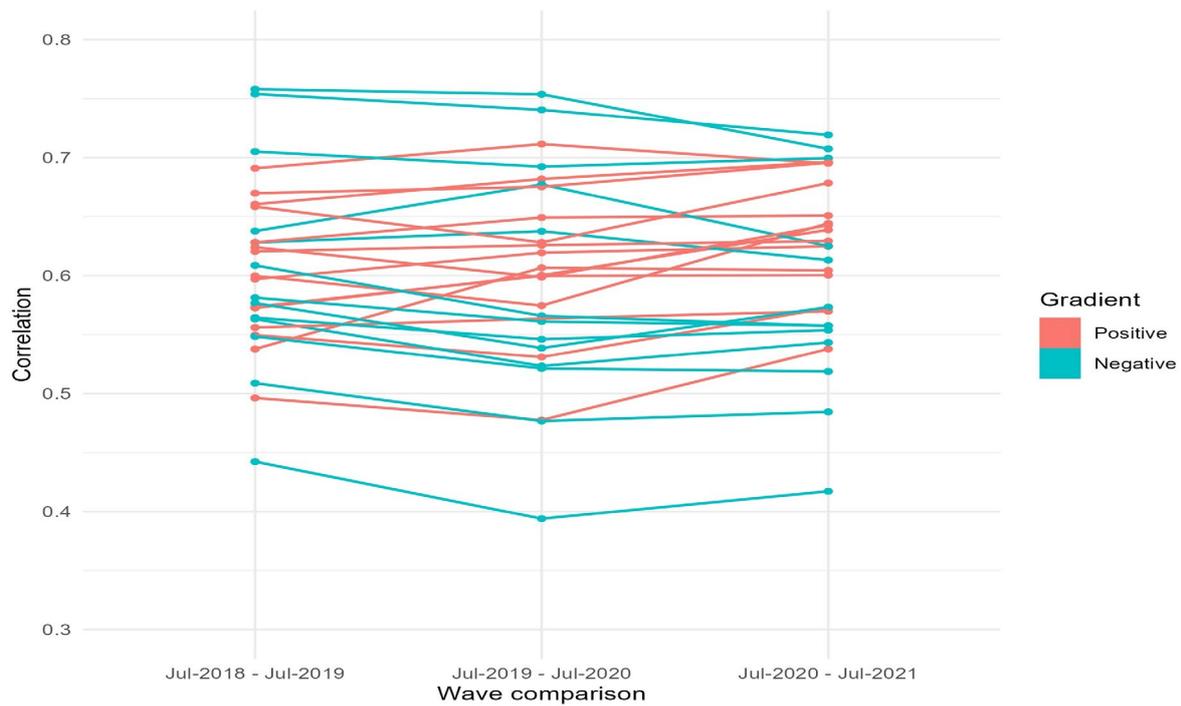


The p-values represent significance of direction of the estimated gradient, either positive or negative.

The changes in the inter-wave consistency of 28 additional attitudinal items are shown in Figure 5 (with the results of statistical significance testing shown in Appendix 4).

For the 28 individual attitudinal items, 15 have positive gradients (two of which are significant) and 13 have negative gradients (three of which are significant).

Figure 5 Inter-wave correlations: Inter-wave consistency of the 28 items over time



To the extent that just over one-half of the items investigated show some increase in consistency there may be a tendency to conclude that this provides some support for the CST hypothesis, however, other possible explanations cannot be ruled out.

One such explanation is ‘freezing’ (Sturgis et al., 2009, 278). Freezing can result in respondent-related measurement error if respondents answer questions that are repeated across waves in a consistent (invariant) fashion from wave to wave,

rather than cognitively updating and accurately reporting their responses to reflect possible changes in their attitudes and opinions from wave to wave.

We investigate this by calculating the number of questions for which each respondent showed no variation at across the four waves. Table 3 shows that, on average, respondents gave the same response across all four waves to 10.7 questions (out of 28) (median=11; mode=11; Range=0–22).

Table 2 The number of questions for which a respondent gave the same response in all four waves

Items	Frequency (count)	Per cent (%)
0	2	0.2
1	3	0.2
2	14	1.1
3	23	1.7
4	18	1.4
5	40	3.0
6	65	4.9
7	98	7.4
8	100	7.5
9	127	9.5
10	134	10.1
11	144	10.8
12	138	10.4
13	133	10.0
14	101	7.6
15	73	5.5
16	59	4.4
17	26	2.0
18	17	1.3
19	10	0.8
21	4	0.3
22	1	0.1
Mean	10.7	
Mode	11	
Median	11	

Of course, to some extent, this level of response invariance may be due to the innate stability of the attitudes being measured, as seems plausible for some items. But to look further into this pattern of responding, we looked at differences in the level of response invariance when limiting our analysis to the 24 items that measure attitudes using a 5-point scale.

(Doing this enables a comparison of response invariance to questions using the same scale format for each question and over time.)

The results of this analysis, shown in Table 3, indicate that the mean number of times a response code of 1 is given to a 5-point scale question across all four

waves is 1.21. Code 2, which relates to the second strongest level of agreement, is the most invariant (i.e., it is provided as a consistent response across all four waves to an average of 3.81 questions). The mean number of times the response codes 3, 4, and 5 are selected without variation across all four waves are 0.62, 1.48 and 0.68, respectively. In percentage terms, 92 per cent of panellists respond with response code of 2 across all four waves at least once. This is followed by

response code 4 (72%), response code 1 (61%), response code 3 (39%) and response code 5 (36%).

A plausible, and we think likely, explanation for this is that it is a form of satisficing whereby some respondents, instead of engaging in the full cognitive response process for each question in each wave, instead just opt to 'agree' for some questions.

Table 3 The number of times a panellist gave the same response to each 5-point Likert scale question (24 items) in all four waves

	Response code 1(a) always provided	Response code 2(b) always provided	Response code 3(c) always provided	Response code 4(d) always provided	Response code 5(e) always provided
N	1,333	1,333	1,333	1,333	1,333
Mean	1.21	3.81	0.62	1.48	0.68
Median	1	3	0	1	0
Mode	0	2	0	1	0
Minimum	0	0	0	0	0
Maximum	8	13	4	7	7
Selected 1 or more times (%)	61%	92%	39%	72%	36%

Notes:

- a) Strongly agree/Very positive/Very happy/Very satisfied/A great deal
- b) Agree/Somewhat positive/Happy/ Satisfied/Somewhat
- c) Neither (not always explicit)
- d) Disagree/Negative/Unhappy/Dissatisfied/Only slightly
- e) Strongly disagree/Very negative/Very unhappy/Very dissatisfied/Not at all

These analyses show that the increase in response consistency observed for 15 of the 28 items studied, cannot necessarily be assumed to arise from the opportunity to 'reflect' on and improve previous answers and thereby be seen as evidence of beneficial panel conditioning. As we have shown, other explanations such as

any innate invariability in the attitudes being measured, as well as potentially harmful panel conditioning effects such as inter-wave freezing and satisficing may also be contributing factors.

4.3 Changes in SDR over time

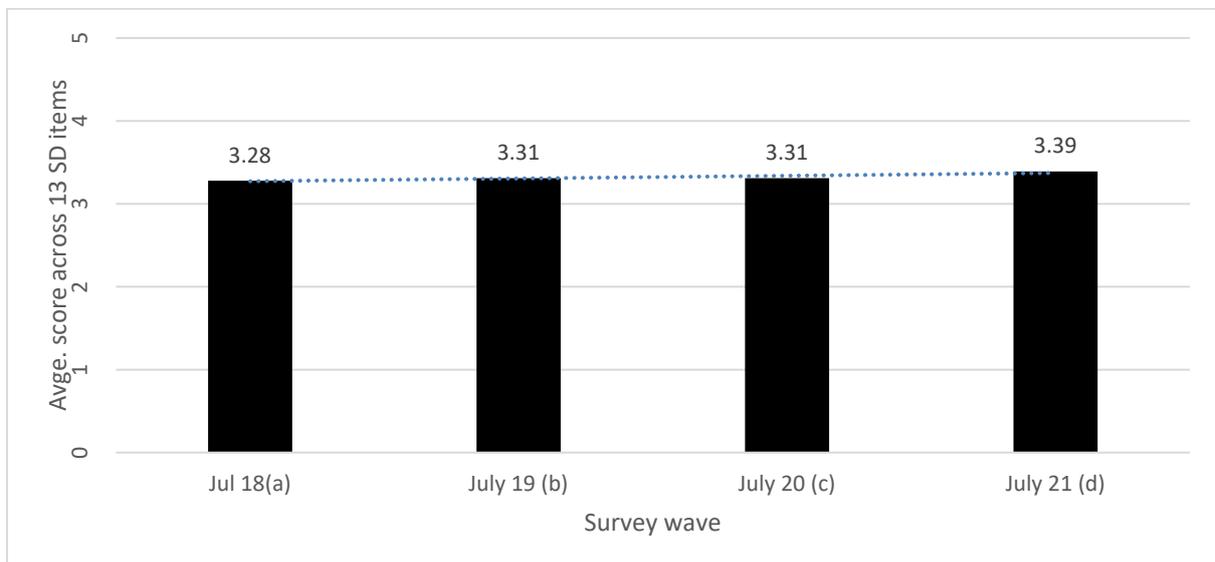
Our third Research Question investigates whether there is any overall change¹⁸ in SDR over time, and if so, does it increase, decrease, or remain static. The research literature on this topic produces mixed results.

Figure 6 shows the aggregated mean scores for the 13 SD items for each of the four waves. Twelve of the 13 items are measured on a five-point-scale and one, originally reported on a three-point-scale, is converted to a binary response (see Appendix 2 for all items). We calculate an additive scale score for each panellist using the following reasoning: 62 scale

points are measured across these 13 items (12*5) + (1*2). Dividing this additive scale score by 13 then results in a minimum possible mean scale score of 1 and a maximum possible mean score of 4.76 (62/13). The direction of the scales used in these questions are standardised so that a score of 1 relates to the least socially desirable response option and a score of 5 the most socially desirable option.

On this basis, the response pattern observed in Figure 6, which shows an *increase* in the aggregated mean score for these 13 items from 3.28 in Wave 1 (July 2018) to 3.39 in Wave 4 (July 2021), reflects an overall significant *increase*, albeit a small one, in SDR over time. This uplift occurs in Wave 4.

Figure 6 Means score for the 13 social desirability (SD) variables per wave



* Results statistically significantly different from 2018 using a Paired Values t-Test ($p < .01$)

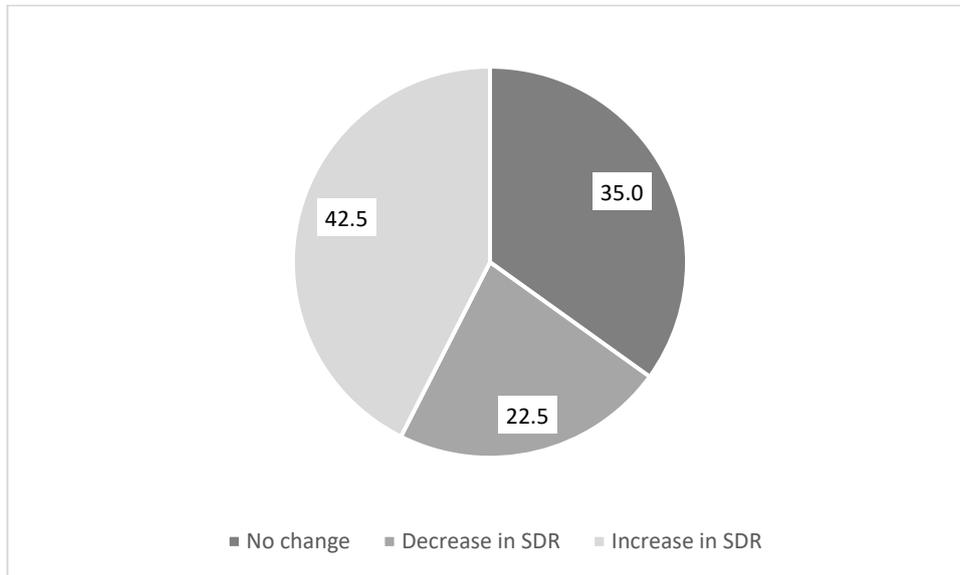
¹⁸ Of note, some panellists may show a decrease in SDR whereas others may show an increase, and still others, no change. Looking only at

overall change masks these individual level changes.

Figure 7 shows, in net terms for these 13 items, the proportion of panellists who displayed an increase in SDR (42.5%), a decrease in SDR (22.5%) or no net

change (35.0%) between Wave 1 and Wave 4. So, again, a pattern whereby more panellists are showing an increase in SDR than a decrease in SDR.

Figure 7 Proportion of panellists who showed a net increase, a decrease, or no change in social desirability (SD) across the 13 SD items



Another way of exploring changes in the patterns of response to these designated SD items is to count the number of items for which panellists showed an increase or decrease in SDR between Waves 1 and 4. The mean number of items of items moved in either direction for selected subgroups of panellists is shown in Table 4.

On average, panellists recorded an increase in SDR in 2021 compared to 2018 for 1.25 questions and a decrease in SDR for an average of 0.70 questions. A resultant net increase in SDR of 0.55 items, from the 13 items examined.

Subgroup analyses reveals that the panellists who show the largest increase in SDR across this set of items between Wave 1 and Wave 4 are Conservative voters (1.65 items), those living in the most disadvantaged areas (1.61), those

with a highest level of educational attainment of less than Year 12 (1.57 items), and those in relatively low-income households (1.50).

Concurrent with the above, panellists also showed a decrease in SDR across this range of SD items. Those panellists who showed the largest decrease in SDR, compared to an overall mean of 0.70 items, are those with a less than Year 12 level of educational attainment (1.0), Indigenous panellists (0.93), those living in the most disadvantaged areas (0.91), and Conservative voters (0.90).

All the subgroups examined showed a net increase in SDR ranging from a mean increase of 0.21 items for those with religious beliefs to 0.87 for those aged 75 years and over.

Table 4 Mean number of variables for which there was an increase or decrease in socially desirable responding

Subgroup	Mean number of variables		
	Increase in SDR	Decrease in SDR	Net change
Total	1.25	0.7	0.55
Male	1.26	0.71	0.55
Female	1.24	0.7	0.54
18–34 years	1.2	0.69	0.51
75+ years	1.5	0.63	0.87
Less than Year 12	1.57**	1.00**	0.57
Graduate/Postgraduate	0.98	0.61	0.37
Overseas born, NESB	1.4	0.92	0.48
Not Remote ^a	1.25	0.69	0.56
Remote ^a	1.31	0.86	0.45
Homeowner	1.24	0.68	0.56
Not homeowner	1.25	0.77	0.48
Not daily smoker / Non-smoker	1.23	0.69	0.54
Daily smoker	1.48	0.82	0.66
Employed	1.16	0.71	0.45
Not employed	1.36	0.71	0.65
Religious	0.96	0.75**	0.21
No religion	1.45**	0.64	0.81
Married	1.28	0.72	0.56
Not currently married	1.19	0.68	0.51
Conservative voter at last election ^b	1.65*	0.90	0.75
Progressive voter at last election ^c	0.91	0.52	0.39
Q1 – Low psychological distress quartile ^d	1.4	0.77	0.63
Q4 – High psychological distress quartile ^d	1.21	0.71	0.5
Indigenous	1.18	0.93	0.25
Non-Indigenous	1.25	0.71	0.54
Q1 – Low Income	1.50*	0.74	0.76
Q4 – High Income	1.09	0.7	0.39
Q1 – Most disadvantaged area ^e	1.61**	0.91**	0.70
Q5 – Least disadvantage area ^e	1.11	0.56	0.55

Notes:

*Statistically less than $p < .05$, **Statistically less than $p < .001$. Testing conducted in SPSS Survey Reporter software.

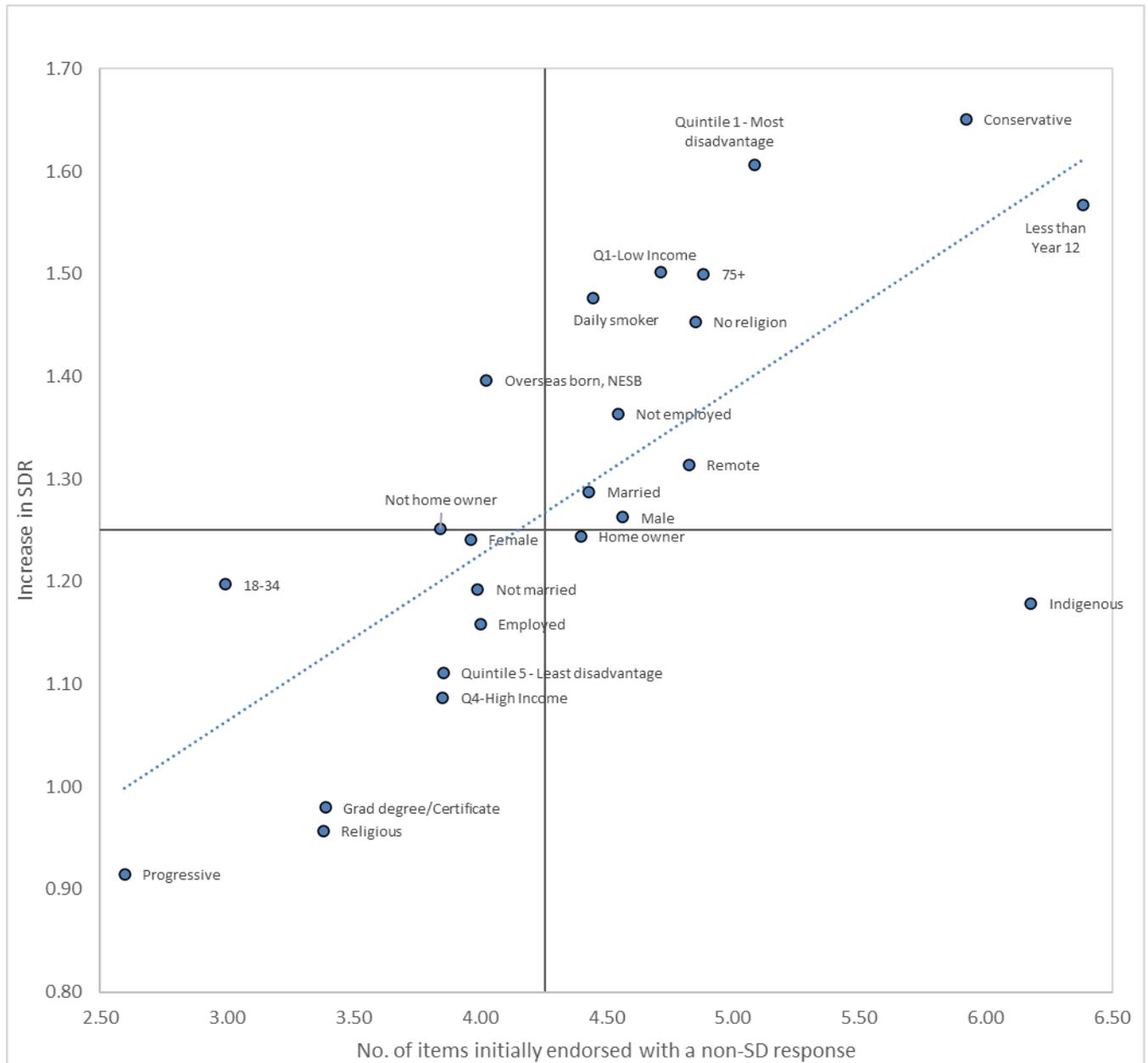
- a. Based on the Australian Bureau of Statistics (ABS) Remoteness Structure (Remote = Remote/Very Remote; Not Remote=Major Cities, Inner Regional and Outer Regional. See <https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/remoteness-structure/remoteness-areas>
- b. Liberal-National Coalition, United Australia Party, Pauline Hanson's One Nation Party.
- c. Australian Labor Party, The Greens.
- d. Refer to Krynen A, Osborne D, Duck IM, Houkamau CA & Sibley CG (2013). Measuring psychological distress in New Zealand: Item response properties and demographic differences in the Kessler-6 screening measure. *New Zealand Journal of Psychology*, 42(2):69–83.
- e. Based on the ABS Index of Relative Socioeconomic Disadvantage. See <https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2033.0.55.001~2016~Main%20Features~IRSD~19>

Figure 8 shows the relationship between the extent to which panellists responded in a non-socially desirable fashion to begin with (in 2018) and subsequent movement in a socially desirable direction (in 2021). The average number of items to receive a non-SD response to begin with is 4.26 and the average number of items moved in a socially desirable direction is 1.25 (as shown by the gridlines). The x-axis shows the number of questions answered in 2018 with a non-SD response. The y-axis shows the number of items for which there was a subsequent increase in SDR between 2018 and 2021. The trendline shows that when more items are initially endorsed with a non-SD response the subsequent increase in SDR tends to be larger. Two groups of panellists stand out. The first group is those in the bottom left quadrant of Figure 8 who endorsed both a relatively low number of items with a non-SD response to begin with (i.e., a relatively high number of SD responses) and showed a relatively small increase in SDR

(as their capacity to move in this direction is somewhat constrained). Included amongst this group are Progressive voters, 18–34-year-olds, graduates, and those who identify as religious. The second main group is in the upper right quadrant which contains those who, from a starting point of endorsing a relatively high number of items with a non-SD response rebound somewhat by endorsing relatively more SD options over time. Most conspicuous among this group are those who have not completed secondary school, Conservative voters, and those residing in areas of relatively high socioeconomic disadvantage.

This analysis shows that the number of items initially endorsed with a non-SD response (or a SD response) is an important factor to consider when examining changes in SDR over time. The subgroup differences apparent in Figure 8 suggest a differential impact which warrants further investigation.

Figure 8 The relationship between a non-social desirability (SD) start point and increases in SDR over time



Base: Panellists who moved in a socially desirable direction between 2018 and 2021 (n=686)

In order to identify any underlying factors that might predict changes in SDR over time, a multinomial logistic regression was

conducted entering predictors using a backwards stepwise method.¹⁹ The dependent variable contained three

¹⁹ The backward stepwise approach with an exclusion criteria of p=.01 was chosen given the exploratory nature of this research where

our interest is in identifying possible relationships between our predictors and changes in SDR.

mutually exclusive outcomes (a net increase in SDR, a net decrease in SDR and no net change). Appendix 3 shows the predictors entered and removed from the model and all model outputs.

The fit between the model improved with the removal of non-significant predictor variables (likelihood ratio $\chi^2(22) = 210.642$, $p < .001$, Nagelkerke pseudo- $R^2 = .184$). The model correctly classified 55.2 per cent of cases overall and was most effective in classifying those who showed an increase in SDR (70.2% of such cases being correctly predicted by the model).

Those variables that best predicted an increase in SDR are not having a religion ($B = .315$, $SE = .155$, $OR = 1.371$, $p < .05$), and being aged 18–34 years ($B = .529$, $SE = .236$, $OR = 1.697$, $p < .05$). In addition, for each increase in the number of questions initially endorsed with a non-socially desirable response in 2018, the odds of showing an increase in SDR increase by 32 per cent ($B = .279$, $SE = .086$, $OR = 1.322$, $p < .001$). The socioeconomic status of the area in which one lives approaches significance as a predictor of an increase in SDR ($B = .092$, $SE = .052$, $OR = 1.096$, $p = .079$).

The characteristic that best predicts a decrease in SDR is not being politically active ($B = -.479$, $SE = .181$, $OR = 1.615$, $p < .01$). Additionally, for each 1-point increase in the conscientiousness scale (as measured by the short-form Big 5 Personality Inventory (see Rammstedt & John, 2007) the prospects of belonging to the decrease in SDR group increases by 19 per cent ($B = .176$, $SE = .088$, $OR = 1.192$, $p < .05$). The socioeconomic status of the area in which one lives is also a predictor of a decrease in SDR, as locational disadvantage decreases the likelihood of being in the decrease in SDR group

increases ($B = 0.117$, $SE = .059$, $OR = 1.124$, $p < .05$). Also, for each increase in the number of questions initially endorsed with a non-SD response in 2018, the odds of a decrease in SDR increased by about 10 per cent ($B = .092$, $SE = .030$, $OR = 1.096$, $p < .01$). The personality trait of Agreeableness (as measured by the short-form Big 5 Personality Inventory (see Rammstedt & John, 2007) is associated with a reduced likelihood of belonging to the decrease in SDR group ($B = -.214$, $SE = .086$, $OR = .807$, $p < .05$). The personality trait of Openness (as measured by the short-form Big 5 Personality Inventory (see Rammstedt & John, 2007) approaches statistical significance as a factor predicting a reduced likelihood of belonging to the decrease in SDR group ($B = -.139$, $SE = .082$, $OR = .870$, $p = .089$).

When all other variables are held constant, the factors that contribute most to a change in SDR over time are being aged 18–34 years, whether or not someone is politically active/engaged, having or not having a religion, the socioeconomic profile of the area where one lives, the personality traits of conscientiousness and agreeableness and the extent to which non-SD responses are endorsed to begin with.

With the model having a predictive accuracy of 55 per cent and explaining just 18 per cent of the variance in our dependent variable clearly other factors, not captured by the model, contribute to changes in the level of SDR over time (this is addressed in the following Discussion section).

5 Discussion

5.1 Our Research Questions

First, we find only weak support for the CST hypothesis that panel conditioning leads to an increase in the validity of attitudes.

The fact the validity of responses over time increased significantly for two of the five constructs measured could be seen as in accord with the CST hypotheses that, for some items at least, the opportunity to 'reflect' on previous answers leads to the adoption of more considered opinions to those given when first responding to a question. If this trait is indeed attributable to the 'reflection' cognitive response process (see Figure 2), then this could be regarded as a beneficial effect of panel conditioning for at least some items (Amaya et al., 2021; Kraemer et al., 2023; Struminskaya & Bosnjak, 2021; and Sturgis et al., 2009 all posit that some panel conditioning effects are beneficial in that they lead to more considered opinions being reported by panellists).

With respect to our second Research Question, we do not find clear evidence of an increase in response consistency over time and, as such, the second tenet of the CST is not supported. In fact, for the 33 items measured over these four waves, panellists are almost as likely to show a decrease in response consistency as they are an increase.

Third, alongside the above, 43 per cent of panellists show an increase in SDR in Wave 4 compared to Wave 1, almost one-quarter (23%) show a decrease in SDR and about a one-third (35%) show no net change. As such, the CST-based expectation that we would observe an

overall decrease in SDR over time is not borne out by these data.

Our finding that most respondents did not, on balance, report less socially desirable responses, is consistent with another recent study on a probability-based mixed mode research panel. In their investigation into panel conditioning on the GESIS panel, Kraemer et al. (2023) also hypothesised that panel conditioning would lead to a reduction in SDR but, like us, found the opposite. In the GESIS study the measure used to detect changes in SDR was to compare item non-response rates to 23 sensitive questions asked of two cohorts of panellists being exposed to these questions for either the first or second time (12-months hence). The expectation was that panel conditioning would result in the repeat respondents showing lower levels of item non-response to the sensitive questions. They found the opposite, 'indicating that respondents with higher experience levels may have been more prone to SDR than low experienced respondents' (Kraemer et al., 2023, 19).

While it is not possible to disentangle whether the increases in SDR observed in our study and in the GESIS study are due to actual changes in attitudes or due to measurement error arising from panel conditioning, the findings from these two studies do raise the prospect that SD may manifest in a different way than originally hypothesised.

So, what might these other factors be, and if the CST does not explain changes in the level of panel-induced SDR over time, what does? From a theoretical perspective a plausible explanation could be that a panellist's ongoing active participation in a research panel and repeated intermittent exposure to the same questions may constitute the development of a familiar 'relationship' between the research panel and the panellist. This relationship might be further fostered by common panel management activities such as expressions of gratitude from the panel operator to panellists for participating in the panel, the payment of incentives, reminder activity, routine panel management activities such as updating profiling information and the sharing of results and newsletters. Many of these activities are aimed at creating an identity for the research panel and a sense that a panellist is contributing to, and belongs to, something worthwhile. Under such a scenario it seems plausible that some panellists would be less inclined to admit to attitudes which violate social norms and more inclined to provide socially desirable responses motivated by their 1) need for social approval, 2) self-presentation concerns and/or 3) impression management. If so, under such conditions, according to Krumpal's explanation of the three mechanisms driving SD, the development of a relationship between the research panel and a panellist would tend to 'yield (more) socially desirable responses on the individual level and [thereby] a predictable bias in survey estimates on the aggregate level' (Krumpal, 2013, 2042).

We hope to undertake further research to test this theoretical explanation.

Finally, while at face-value it is unsurprising that the number of questions initially endorsed with a socially desirable or socially undesirable response effectively prescribes the amount of subsequent movement in SDR, the fact that the impact of previous responses on subsequent changes in SDR varies considerably across subgroups, a finding that is masked if changes in SD are only examined at a total sample level, does suggest a range of factors are contributing to the observed changes in SDR. The latent characteristics found in this study to mediate changes in SDR over time are sociodemographic characteristics such as the age of panel members and the socioeconomic profile of the area in which they live, having or not having religious beliefs, one's level of political activity/engagement, and personality traits related to conscientiousness and agreeableness.

5.2 Strengths and limitations

The reliance of this study upon a single cohort of panellists who completed these four waves of the survey is both a strength and a weakness of our approach. The fact that the panellists in our sample completed all four waves is a strength in that it eliminates the possible confounding effects of unit non-response and panel attrition, but the fact that our sample is effectively a single cohort of Life in Australia™ panellists is a weakness because this design does not support a comparison of panellists with enough different levels of panel experience.

This study relies on identifying or deriving indicators of panel conditioning by conducting a secondary analysis of existing data and comparing the results against theoretical expectations, rather than using purposely-designed data. This means, for example, that due to the way

that ‘Don’t Know’ answers are displayed in these questionnaires,²⁰ that we could not use changes in the level of item non-response, as others have, to measure aspects of attitudinal stabilisation (cf. Sturgis et al., 2009) or SD (cf. Kraemer et al., 2023). The strongest assertion that can be made from our design is whether we observe patterns that are consistent (or not) with our *a priori* panel conditioning hypotheses. We cannot definitively rule out other causes for the changes in the response patterns that we have observed, including actual changes in attitudes over time and the usual variations in how respondents’ map their opinions to rating scales.

Lastly, it is almost certainly true that no two research panels are the same. The probability online panels that are now more rapidly emerging around the world are certainly very different from the large-scale longitudinal household surveys which spawned a lot of the initial research into panel conditioning, quite different from each other, and very different from the vast number of non-probability research panels. Not only are there differences in how such panels are sampled, recruited, have data collected from panellists, and are maintained but there are many other differences such as wave cadence, incentive structures, the general tone of the panels, the vastly

different topics measured, differences in questionnaire design practices/formats (e.g., small screen optimisation) and so on that may impact the generalisability of these findings to other probability panel settings.

The Life in Australia™ panel is subject to the errors of representation and measurement common to probability-based research panels (see Lavrakas et al., 2020 for a full discussion). Typical of such panels, Life in Australia™ covers both the online and offline population.

Questionnaires are administered to the offline population, which comprises less than five per cent of the panel members, via interviewer-administered telephone surveys (CATI) whereas the vast bulk of questionnaires (>95%) are administered to online panellists via a self-administered online mode of data collection (CAWI). To the extent that these two modes differ in terms of the amount of mode-related SD bias they induce, the results from this study may not be generalisable to other probability-based online panels, particularly those that rely solely on a CAWI mode of data collection.

²⁰ Given that the Mapping Social Cohesion Survey started out as a CATI survey when transitioning to Life in Australia™ it was decided Don’t Know/ Can’t Say and Refused options would not be provided on the same page as the question stem and substantive response options (as they are typically not

read out by an interviewer in a CATI survey) but instead only displayed to respondents if they tried to advance to the next item without selecting a substantive response. When this happened, the respondents were required to select either a Don’t Know or Prefer Not to Say option to advance in the questionnaire.

6 Concluding remarks and next steps

This study contributes to the emerging body of knowledge about panel conditioning effects in probability-based online panels and to the discussion as to whether such effects might be deemed beneficial or detrimental to data quality. Our study finds insufficient evidence to support of the assertion that panel conditioning leads to an overall increase in the validity of attitudinal reporting. Also, when examining responses across four equidistant annual waves of this survey program we find no convincing evidence of increased response consistency. Furthermore, the extent to which some items demonstrate a greater level of inter-wave consistency than others cannot automatically be assumed to be a beneficial effect of panel conditioning. As we have shown, other explanations remain plausible such as the innate invariability in the attitudes being measured, as well as potentially harmful panel conditioning effects such as inter-wave freezing and satisficing.

As to whether being on a research panel for a period increases or decreases the tendency for panellists to report socially desirable attitudes, while we find

movement in both directions mitigated by factors such as age, the socioeconomic status of one's place of residence, personality traits and involvement in political activity. Overall, we observe a net increase in SDR. Fundamentally we find a panellist's starting out point in answering a set of items that are prone to SD bias (i.e., the number of items initially endorsed in a socially desirable or undesirable fashion), is a strong predictor of the quantum of change in SDR.

The focus of this paper is whether there is evidence that panel conditioning effects the reporting of attitudes in a manner consistent with the CST hypotheses. This is only one line of inquiry. As noted by Struminskaya and Bosnjak (2021, 273) there are 'three types of changes that can result from panel conditioning i) changes in reporting behaviour; (ii) changes in actual behaviour, attitudes, or knowledge; and (iii) a combination of actual changes and changes in reporting.' Thus, there is more research we would like to do in this area. Topics for further investigation include but are not limited to the further exploration of the topics covered in this paper, as well as investigating both the beneficial and detrimental effects of panel conditioning on panellists' questionnaire completion behaviour through measures designed to capture evidence of changes in the quality of the data provided by panellists over time.

References

- Amaya A, Hatley N & Lau A (2021). Measuring the Risks of Panel Conditioning in Survey Research. Pew Research Center, <https://www.pewresearch.org/methods/2021/06/09/measuring-the-risks-of-panel-conditioning-in-survey-research/> (accessed 6 July 2023)
- Anand S (2008). Satisficing. In PJ. Lavrakas (ed.) *Encyclopedia of survey research methods*, Vol. 2, Sage, 797–799.
- Bailar BA (1975). Effects of Rotation Group Bias on Estimates from Panel Surveys. *Journal of the American Statistical Association* 70:23–30. <https://www.tandfonline.com/doi/abs/10.1080/01621459.1975.10480255>
- Brannen J (1993). Research Notes: The effects of research on participants: Findings from a study of mothers and employment. *The Sociological Review*, 41(2):328–346. <https://doi.org/10.1111/j.1467-954X.1993.tb00068.x>
- Broedling LA (1974). On more reliably employing the concept of ‘reliability’. *Public Opinion Quarterly* 38(3):372–378.
- Callegaro M (2008) Social desirability. In PJ Lavrakas (ed.) *Encyclopedia of survey research methods*, Vol. 2, Sage, 825–826.
- Cantwell PJ (2008). Panel conditioning. In PJ Lavrakas (ed.) *Encyclopedia of survey research methods*, Vol. 2, Sage, 567.
- Clinton JD (2001). Panel Bias from Attrition and Conditioning: A Case Study of the Knowledge Networks Panel. Accessed 6 July 2023 https://www.researchgate.net/publication/238292479_Panel_Bias_from_Attrition_and_Conditioning_A_Case_Study_of_the_Knowledge_Networks_Panel
- Davison AC & Hinkley DV (1997). *Bootstrap methods and their applications*. Cambridge University Press.
- European Society for Opinion and Market Research (ESOMAR) (2022). *Global Market Research report*, ESOMAR, <https://esomar.org/newsroom/esomar-publishes-the-2022-global-market-research-report>
- Halpern-Manners A & Warren JR (2012). Panel conditioning in longitudinal studies: Evidence from labor force items in the Current Population Survey. *Demography* 49:1499–1519. <https://read.dukeupress.edu/demography/article/49/4/1499/169764/Panel-Conditioning-in-Longitudinal-Studies>
- Halpern-Manners A, Warren JR & Torche F (2017). Panel conditioning in the General Social Survey. *Sociological Methods & Research* 46:103–124. <https://journals.sagepub.com/doi/abs/10.1177/0049124114532445>

- Kaczmirek L, Phillips B, Pennay D & Neiger N (2019). *Building a probability-based online panel: Life in Australia™*. CSRM & SRC Methods Paper No. 2/2019. Centre for Social Research & Methods, Research School of Social Sciences, College of Arts & Social Sciences, ANU. <https://srcentre.com.au/our-research/methods-research/Building%20a%20probability-based%20online%20panel-Life%20in%20Australia%20-%202.0.pdf> (accessed 3 March 2023).
- Kennedy C, Popky D & Keeter S (2023). *How polling has changed in the 21st Century*. Pew Research Center, <https://www.pewresearch.org/methods/2023/04/19/how-public-polling-has-changed-in-the-21st-century/> (accessed on 9 May 2023)
- Kraemer F, Silber H, Struminskaya B, Bosnjak M, Koßmann J & Weiß B (2023). Panel conditioning in a German probability-based longitudinal study: A comparison of respondents with different levels of survey experience. *Journal of Survey Statistics and Methodology*. <https://doi.org/10.1093/jssam/smad004>.
- Krumpal I (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity* 47:2025–2047. <https://doi.org/10.1007/s11135-011-9640-9>
- Lavrakas PJ, Pennay D, Neiger D & Phillips B (2022). Comparing probability-based surveys and nonprobability online panel surveys in Australia: A total survey error perspective. *Survey Research Methods* 16(2):241–266. <https://doi.org/10.18148/srm/2022.v16i2.7907>
- McCormick MK, Butler DM & Singh RP (1992). Investigating time in sample effect for the Survey of Income and Program Participation. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 554–559, Alexandria, VA.
- Mavletova A & Lynn P (2019). Item nonresponse rates and panel conditioning in a longitudinal survey among youth. *Field Methods*, 31(2):95–115. <https://doi.org/10.1177/1525822X18825103>.
- Pennell SG & Lepkowski JN (1992). Panel conditioning effects in the Survey of Income and Program Participation. In *Proceedings of the Survey Research Methods Section of the American Statistical Association*, Vol. 566571. Alexandria, VA.
- Pevalin DJ (2000). Multiple applications of the GHQ-12 in a general population sample: An investigation of long-term retest effects. *Social Psychiatry and Psychiatric Epidemiology* 35 (11):508–512. <https://www.proquest.com/docview/236136200/2C354CDD391E4A80PQ/4?accountid=8330>

- Phillips B, VanderGert D, Myers P, Neiger D & Challice G (2022). Recruiting Life in Australia™ using A-BS, IVR and SMS Push-to-Web. Paper presented at the 8th ACSPRI Social Science Methodology Conference, Online, 24 November.
- Rammstedt B & John OP (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality* 41(1):203–212.
<https://doi.org/10.1016/j.jrp.2006.02.001>
- Struminskaya B & Bosnjak M (2021). Panel conditioning: Types, causes, and empirical evidence of what we know so far. In Lynn P (ed.) *Advances in longitudinal survey methodology*, 272–301. John Wiley and Sons. <https://doi.org/10.1002/9781119376965.ch12>
- Sturgis P, Allum N & Brunton-Smith I (2009). Attitudes over time: The psychology of panel conditioning. In Lynn P (ed.) *Methodology of longitudinal surveys*, 113–126. John Wiley and Sons.
<https://ebookcentral-proquest-com.virtual.anu.edu.au/lib/anu/reader.action?docID=416458&ppg=3>
- Torche F, Warren JR, Halpern-Manners A & Valenzuela E (2012). Panel conditioning in a longitudinal study of adolescents' substance use: Evidence from an experiment. *Social Forces* 90:891–918.
<https://doi.org/10.1093/sf/sor006>
- Tourangeau R, Rips LJ & Rasinski K (2000). *The psychology of survey response*. Cambridge University Press
- Waterton J & Lievesley D (1989). Evidence of conditioning effects in the British social attitudes panel. In Kasprzyk D, Duncan G, Kalton G & Singh MP (eds) *Panel surveys*, 319–339. John Wiley and Sons.
- Weir, DR and Smith, JP (2007). Do panel surveys really make people sick? A commentary on Wilson and Howell (60: 11, 2005, 2623–2627). *Social Science & Medicine* 65(6): 1071–1077.
- Wilson, SE and Howell, BL (2005). Do panel surveys make people sick? US arthritis trends in the health and retirement survey. *Social Science and Medicine* 60: 2623–2627

Appendix 1 – Comparative profile of panellists

Table 5 Comparative profile of panellists by selected characteristics: Participated in fewer than four waves and participated in all four waves

Selected characteristics	Pct Participated in fewer than 4 waves (N=3,186) (a)	Pct Participated in all 4 waves (N=1,333) (b)
Male	41.7	48.5**
Female	57.8**	51.3
18–34 years	22.3**	10.7
35–44 years	17.6**	13.1
45–54 years	16.4	16.8
55–64 years	17.2	21.3**
65–74 years	16.8	24.2**
75+ years	9.4	14**
Post graduate qualifications	19.2	25.8**
Bachelors or Adv. Diploma	29.7	33.9*
Certif III & IV to Year 12	24.1	24.5
Less than Year 12	9.7	12.2*
Not Remote area	92	92.1
Remote area	7.7	7.9
NESB country of birth	15.7**	11.2
ESB country of birth	10.3	15**
Owns home outright or with a mortgage	64.9	74.1**
Daily smoker	7.5	8
Employed	46.9	55**
No religion	62.3	59.7
Religion	37.7	40.3
Currently married	42.1	56.9**
Never married	24.5*	20.8
Internet frequency, never	0.8	2.8*
Less often	4.3	5.8
At least daily	80**	91.1
Social media frequency, never	0	0

Selected characteristics	Pct Participated in fewer than 4 waves (N=3,186) (a)	Pct Participated in all 4 waves (N=1,333) (b)
Less often	41.3*	44.2
At least daily	44.1	55.4*
Voted for a Progressive party at the last federal election	40.6	42.1
Voted for a Conservative party at the last federal election	32.5	37.4**
Self-reported Psychological Distress - 1-Low	41.7**	20.3
-2-	22.8	26.9**
-3-	16.1	17.6
4-High	19.3	35.3**
Non-Indigenous	98.1	97.8
Indigenous	1.6	2.1
Q1-Low Income	18.1	19.4
Q2	22	25.1
Q3	11.4	12.5
Q4-High Income	29.2	36.3

Difference between columns a) and b). *p<.05,** p <.01

Appendix 2 – Substantive variables

Table 6 List of substantive variables and analytic use

Item	Scale	Analytical use		
		Validity	Internal Consistency	Social Desirability
People on low incomes get enough gov financial support	1-Strongly Agree, 5-Strongly Disagree	x	x	x
Gap between high and low incomes too large	1-Strongly Agree, 5-Strongly Disagree		x	x
What do you think of no. of immigrants accepted into Aus	1-Too High, 2 – About right, 3- Too Low		x	x
Accepting immigrants from many different countries makes Aus stronger	1-Strongly Agree, 5-Strongly Disagree	x	x	x
Ethic minorities in Aus SHOULD be given Aus gov support to maintain culture	1-Strongly Agree, 5-Strongly Disagree	x	x	x
Multiculturalism has been good for Aus	1-Strongly Agree, 5-Strongly Disagree		x	x
Agree/disagree to reject people on basis of: race or ethnicity	1-Strongly Agree, 5-Strongly Disagree		x	x
Agree/disagree to reject people on basis of their religion	1-Strongly Agree, 5-Strongly Disagree		x	x
Is your attitude positive, negative, neutral towards: Muslims	1-Strongly Positive, 5-Strongly Negative		x	x
Agree or disagree – Immigrants are generally good for Australia’s economy	1-Strongly Agree, 5-Strongly Disagree		x	x
Agree or disagree – Immigrants take jobs away	1-Strongly Agree, 5-Strongly Disagree		x	x
Agree or disagree – Immigrants improve Australian society by bringing new ideas and cultures	1-Strongly Agree, 5-Strongly Disagree		x	x
Aus is a land of economic opportunity	1-Strongly Agree, 5-Strongly Disagree	x	x	
In general, are you optimistic or pessimistic about Australia’s future?	1-Strongly Optimistic, 5- Strongly Pessimistic		x	
How often can federal gov be trusted to do right thing	1 – Almost always, 4- Almost never		x	
Good/bad way of governing Aus: Strong leader	1 – Very good, 4- Very bad		x	

Item	Scale	Analytical use		
		Validity	Internal Consistency	Social Desirability
The system of government in Aus needs change / is fine	1-Strongly Agree, 5-Strongly Disagree		x	
Extent you take pride in Australian way of life	1-A great deal, 4- Not at all	x	x	
Extent you have sense of belonging	1-A great deal, 4- Not at all	x	x	
Maintaining Australian way of life is important	1-Strongly Agree, 5-Strongly Disagree		x	
Is your attitude positive, negative, neutral towards: Christians	1-Very Positive Agree, 5-Very Negative		x	
Is your attitude positive, negative, neutral towards: Buddhists	1-Very Positive Agree, 5-Very Negative		X	
Growing economic ties between Australia and other countries, sometimes referred to as globalisation (Good/Bad for Australia)	1 – Very good, 4-Very bad		x	
Generally speaking would you say most people can be trusted / can't be too careful	1 – Can be trusted, 2-Can't be too careful		x	
Over the last year have you been happy / unhappy	1-Very happy, 5-Very unhappy	x	x	
In 3 or 4 yrs do you think your life in Aus will be better / worse	1-Very happy, 5-Very unhappy		x	
People in your local area: are willing to help your neighbours	1-Much Improved, 5-Much worse		x	
Your local area is a place where people from different ethnic backgrounds get on well together	1-Strongly Agree, 5-Strongly Disagree		x	
How satisfied with financial situation	1-Very Satisfied, 5-Very Dissatisfied	x		
Last 3 years or so: Joined a boycott of a product or company	1-Yes,2-No	x		
Last 3 years or so: Written or spoken to a Federal or State Member of Parliament	1-Yes,2-No	x		
Are you personally concerned that Australia is too harsh in its treatment of asylum seekers and refugees?	1-A great deal, 5-Not at all			x

Appendix 3 – Modelling

Table 7 Model inputs

Variables	Scale	N	Marginal Percentage
Dependent variable			
Change in SDR	Neutral	420	35.4%
	Decrease in SDR	270	22.7%
	Increase in SDR	497	41.9%
Factors			
Sex [#]	Male	591	49.8%
	Female	596	50.2%
Remoteness Indicator (a) – Remote/Very remote [#]	Not Remote	1098	92.5%
	Remote	89	7.5%
Overseas born, Non-English-speaking background [#]	No	1047	88.2%
	Yes	140	11.8%
Overseas born, English-speaking background [#]	No	1006	84.8%
	Yes	181	15.2%
Resp owns or is buying own home? [#]	No	298	25.1%
	Yes	889	74.9%
Employed [#]	Yes	677	43.0%
	No	510	57.0%
Is Resp a daily smoker? [#]	No	1093	92.1%
	Yes	94	7.9%
Did Resp report "No Religion"?	Yes	677	57.0%
	No	691	58.2%
Is Resp now married? [#]	Yes	496	41.8%
	No	500	42.1%
Has respondent ever married [#]	No	937	78.9%
	Yes	250	21.1%
Frequency of using the internet [#]	Never	25	2.1%
	Less often	63	5.3%
	At least daily	1099	92.6%
Social media frequency [#]	Less often	533	44.9%
	At least daily	654	55.1%
Self-reported Psychological Distress (b) [#]	Low psych distress	425	35.8%
	2	211	17.8%
	3	262	22.1%
	High psych distress	289	24.3%
	Non-progressive	668	56.3%

Variables	Scale	N	Marginal Percentage
Vote choice at the last federal election (c)	Conservative [#]	433	36.5%
University graduate / Postgraduate [#]	No	447	37.7%
	Yes	740	62.3%
Age group	Aged 18 — 34 years	128	10.8%
	Aged 35 years and over	1059	89.2%
Politically active (e)	Not politically active	324	27.3%
	Politically active	863	72.7%
Indigeneity	Non-Indigenous	1164	98.1%
	Indigenous	23	1.9%
Household Income quartiles [#]	Q1-Low Income	229	19.3%
	Q2	320	27.0%
	Q3	161	13.6%
	Q4-High Income	477	40.2%
Covariates			
The number of variables each respondent initially endorsed with the a socially desirable response in 2018 (Wave 1)			
Extraversion – as measured by the Big 5 Personality Inventory. BFI-10 (f)			
Agreeableness – as measured by the Big 5 Personality Inventory. BFI-10 (f) #			
Conscientiousness – as measured by the Big 5 Personality Inventory. BFI-10 (f)			
Neuroticism – as measured by the Big 5 Personality Inventory. BFI-10 (f) #			
Openness – as measured by the Big 5 Personality Inventory. BFI-10 (f)			
The number of variables each respondent initially endorsed with the a non-socially desirable response in 2018 (Wave 1)			
Index of Relative Socioeconomic Disadvantage			
Valid		1188	
Missing		145	
Total		1333	

Notes: # Removed from model via Stepwise Method Backward elimination

- a) Based on the Australian Bureau of Statistics (ABS) Remoteness Structure (Remote = Remote/Very Remote; Not Remote= Major Cities, Inner Regional and Outer Regional. See <https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/remoteness-structure/remoteness-areas>)
- b) Refer to Krynen A, Osborne D, Duck IM, Houkamau CA & Sibley CG (2013). Measuring psychological distress in New Zealand: Item response properties and demographic differences in the Kessler-6 screening measure. *New Zealand Journal of Psychology*, 42(2), 69–83.

- c) Australian Labor Party, The Greens.
- d) Liberal-National Coalition, United Australia Party, Pauline Hanson’s One Nation Party.
- e) Derived variable from 2018 survey – Signed a petition, Written or spoken to a federal Member of Parliament, Joined a boycott of a product of=r company, Attended a demonstration or protest.
- f) Rammstedt B & John OP (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212.
- g) Refer to https://www.rand.org/health-care/surveys_tools/sdrs.html

Overall model evaluation

The fit between the model containing only the intercept data improved with the addition of the predictor variables. Likelihood ratio $\chi^2(22) = 210.642$, $p < .001$. The Nagelkerke pseudo-R² indicates approximately 18 per cent of the variance in SDR patterns is accounted for by the predictors overall.

The model correctly classified 55 per cent of cases overall and 70 per cent of those who showed an increase in SDR between 2018 and 2021.

Table 8 Classification table

Observed	Predicted			Per cent correct
	Decrease in SDR	No change	Increase in SDR	
No change	279	13	128	68.4%
Decrease in SDR	122	27	121	10.0%
Increase in SDR	135	13	349	70.2%
Overall Percentage	45.2%	4.5%	50.4%	55.2%

Table 9 Parameter Estimates – Increase in SDR between Waves 1 and 4

Predictor	B	Std. Error	Wald's χ^2	df	p	Exp(B) (odds ratio)
Intercept	-1.910	0.602	10.068	1	0.002	
No religion	0.315	0.155	4.159	1	0.041	1.371
Not a progressive voter	-0.262	0.165	2.536	1	0.111	-0.769
Aged 18—34 years	0.529	0.236	5.022	1	0.025	1.697
Not politically active	0.116	0.168	0.476	1	0.490	1.123
Non-Aboriginal and/or Torres Strait Islander	0.517	0.564	0.840	1	0.359	1.676
Extraversion	0.210	0.145	2.081	1	0.149	1.233
Conscientiousness	-0.141	0.073	3.737	1	0.053	0.868
Agreeableness	-0.086	0.077	1.237	1	0.266	0.918
Index of Relative Socioeconomic Disadvantage	0.092	0.052	3.095	1	0.079	1.096
Conscientiousness	0.055	0.078	0.508	1	0.476	1.057
Openness	-0.125	0.073	2.905	1	0.088	0.882
The number of variables initially endorsed with the a non-socially desirable response	0.279	0.027	110.983	1	0.000	1.322

Reference category – No change in SDR

Table 10 Parameter Estimates – Decrease in SDR between Waves 1 and 4

	B	Std. Error	Wald's χ^2	df	p.	Exp(B) (odds ratio)
Intercept	-1.035	0.586	3.123	1	0.077	
No religion	0.172	0.171	1.012	1	0.314	1.188
Not a progressive voter	0.224	0.183	1.488	1	0.222	1.251
Aged 18–34 years	0.136	0.272	0.250	1	0.617	1.146
Not politically active	0.479	0.181	7.024	1	0.008	1.615
Non-Aboriginal and/or Torres Strait Islander	-0.482	0.540	0.796	1	0.372	0.618
Extraversion	-0.057	0.161	0.126	1	0.723	0.944
Agreeableness	-0.214	0.086	6.232	1	0.013	0.807
Index of Relative Socioeconomic Disadvantage	0.117	0.059	3.966	1	0.046	1.124
Conscientiousness	0.176	0.088	4.005	1	0.045	1.192
Openness	-0.139	0.082	2.900	1	0.089	0.870
The number of variables initially endorsed with a non-socially desirable response	0.092	0.030	9.229	1	0.002	1.096

Reference category – No change in SDR

Appendix 4 – Statistical Testing of inter-wave consistency

Table 11 Gradients of the inter-wave measures of consistency and statistical significance testing

Construct/ Question	Jul 2018 - Jul 2019	Jul 2019 - Jul 2020	July 2020 - Jul 2021	Grad	MOE.for.Cor1	MOE.for.Cor2	MOE.for.Cor3	MOE.for.Grad	P.val.of.Grad.sign	significance_flag
AR	0.753	0.766	0.782	0.014	0.030	0.027	0.027	0.019	0.064	
BE	0.635	0.628	0.625	-0.005	0.037	0.039	0.038	0.024	0.338	
PA	0.582	0.574	0.569	-0.006	0.044	0.045	0.046	0.029	0.316	
SJ	0.658	0.652	0.699	0.020	0.035	0.037	0.034	0.020	0.023	X
WO	0.652	0.624	0.620	-0.016	0.038	0.038	0.041	0.024	0.098	
a1b	0.600	0.575	0.644	0.022	0.039	0.042	0.038	0.025	0.036	X
a1c	0.597	0.619	0.625	0.014	0.041	0.040	0.040	0.028	0.159	
a1d	0.574	0.600	0.600	0.013	0.043	0.044	0.044	0.026	0.157	
a10	0.496	0.478	0.538	0.021	0.046	0.046	0.044	0.029	0.078	
b6a	0.573	0.600	0.639	0.033	0.040	0.038	0.036	0.024	0.003	X
b9	0.550	0.531	0.573	0.011	0.045	0.046	0.045	0.029	0.211	
b10	0.538	0.607	0.604	0.033	0.044	0.038	0.037	0.025	0.007	X
c1	0.758	0.754	0.707	-0.025	0.029	0.030	0.033	0.019	0.004	X
c2a	0.691	0.712	0.695	0.002	0.034	0.032	0.033	0.021	0.428	
c2b	0.660	0.682	0.696	0.018	0.035	0.035	0.036	0.023	0.063	
c2c	0.705	0.692	0.700	-0.003	0.035	0.034	0.035	0.022	0.408	

Construct/ Question	Jul 2018 - Jul 2019	Jul 2019 - Jul 2020	July 2020 - Jul 2021	Grad	MOE.for.Cor1	MOE.for.Cor2	MOE.for.Cor3	MOE.for.Grad	P.val.of.Grad.sign	significance_flag
c3a	0.624	0.599	0.643	0.009	0.039	0.040	0.039	0.026	0.241	
c3b	0.658	0.628	0.679	0.010	0.038	0.039	0.035	0.024	0.200	
c7	0.556	0.564	0.570	0.007	0.043	0.043	0.041	0.027	0.298	
c8	0.564	0.546	0.554	-0.005	0.046	0.049	0.047	0.029	0.365	
c9	0.621	0.626	0.629	0.004	0.040	0.037	0.039	0.025	0.367	
cn7a	0.754	0.741	0.719	-0.017	0.029	0.029	0.031	0.018	0.031	X
cn7b	0.609	0.566	0.557	-0.026	0.039	0.040	0.043	0.026	0.026	X
cn7c	0.670	0.675	0.696	0.013	0.037	0.036	0.033	0.022	0.121	
c13	0.563	0.523	0.543	-0.010	0.043	0.046	0.044	0.028	0.245	
c16a	0.628	0.638	0.613	-0.007	0.038	0.037	0.040	0.025	0.279	
c16b	0.628	0.649	0.651	0.011	0.044	0.041	0.039	0.027	0.207	
c16c	0.638	0.677	0.625	-0.006	0.038	0.037	0.041	0.025	0.306	
e1	0.581	0.561	0.558	-0.012	0.044	0.044	0.045	0.029	0.214	
e2	0.577	0.539	0.573	-0.002	0.043	0.044	0.047	0.030	0.468	
e3	0.509	0.477	0.485	-0.012	0.047	0.048	0.047	0.030	0.212	
f2a	0.548	0.521	0.519	-0.015	0.048	0.048	0.048	0.032	0.179	
f2b	0.442	0.394	0.417	-0.013	0.053	0.056	0.055	0.034	0.233	